Application of statistical methods to the determination of slope in lidar data

David N. Whiteman

Assumptions made in the analysis of both Raman lidar measurements of aerosol extinction and differential absorption lidar (DIAL) measurements of an absorbing species are tested. Statistical analysis techniques are used to enhance the estimation of aerosol extinction and aerosol extinction error that is usually handled using a linear model. It is determined that the most probable extinction value can differ from that of the linear assumption by up to 10% and that differences larger than 50% can occur in the calculation of extinction error. Ignoring error in the number density alters the calculated extinction by up to 3% and that of extinction error by up to 10%. The preceding results were obtained using the least-squares technique. The least-squares technique assumes that the data being regressed are normally distributed. However, the quantity that is usually regressed in aerosol extinction and DIAL calculations is not normally distributed. A technique is presented that allows the required numerical derivative to be determined by regressing only normally distributed data. The results from this technique are compared with the usual procedure. The same concerns raised here regarding appropriate choice of a model in the context of aerosol extinction calculations should apply to DIAL calculations of absorbing species such as water vapor or ozone as well because the numerical derivative that is required is identical.

OCIS codes: 280.1100, 280.3640, 280.1910.

1. Introduction

In the analysis of both Raman lidar measurements of aerosol extinction[1] and differential absorption lidar (DIAL) measurements of an absorbing species such as water vapor or ozone,[2,3] it is necessary to calculate the derivative of the logarithm of the ratio of two quantities. In the case of aerosol extinction that is considered here, these two quantities are the atmospheric number density and the range-corrected lidar-received power. The evaluation of the derivative is handled by fitting linear and quadratic models to Raman lidar data and using the well-known χ^2 confidence test to assess both the best model and the measurement errors. It is shown that assumptions about the behavior of the data can substantially change the answers that are calculated. The most probable assumption is determined using statistical techniques.

The author is with the NASA Goddard Space Flight Center, Code 924, Greenbelt, Maryland 20771-0001. His e-mail address is david.whiteman@gsfc.nasa.gov.

Received 30 September 1998; revised manuscript received 25 January 1999.

2. Extinction and Derivative Equations

To calculate aerosol extinction, the following equation[1] is used:

$$s_{\mathrm{aer}}(\lambda_L, z) = \frac{\dfrac{\mathrm{d}}{\mathrm{d}z}\left\{\ln\left[\dfrac{N(z)}{z^2 P(z)}\right]\right\} - s_{\mathrm{mol}}(\lambda_L, z) - s_{\mathrm{mol}}(\lambda_R, z)}{1 + \dfrac{\lambda_L}{\lambda_R}}. \tag{1}$$

$s_{\mathrm{aer}}(\lambda_L, z)$ is the aerosol extinction coefficient (e.g., in inverse kilometers) at the laser wavelength, where λ_L and λ_R are the laser and Raman-shifted wavelengths, respectively; $s_{\mathrm{mol}}(\lambda_L, z)$ and $s_{\mathrm{mol}}(\lambda_R, z)$ are the molecular (Rayleigh) extinction coefficients; $N(z)$ is the atmospheric number density; z is the altitude of the measurement; and $P(z)$ is the lidar signal that is due to Raman scattering. Equation (1) is applicable for heights such that $z > z_{\mathrm{min}}$, where z_{min} is the altitude above which the lidar system overlap function is unity. For the data to be presented here, that minimum height is approximately 800 m. The height range over which data are analyzed here is thus chosen to be 1–3 km. For the purposes of this paper, the simplifying assumption that aerosol extinction is inversely proportional to the wavelength has been made.

The procedure commonly used[3] for computing a

numerical derivative such as in Eq. (1) is to assume that $\ln[N(z)/z^2P(z)]$ behaves in a linear fashion over some small range of the data and then use a least-squares regression to derive the best-fit straight line. In other words, the assumption is made that, over the range of points being regressed, $\ln[N(z)/z^2P(z)]$ is best approximated by a straight line $a + bz$. A regression is performed and the value of b is reported as the desired derivative, and the standard error in determining b is used as the error in the derivative term in Eq. (1). (This approach will be referred to as the standard technique.) To perform the weighted regression, error must be assigned to both the atmospheric number density $N(z)$ (which is derived from radiosonde measurements) and the range-corrected lidar data $z^2P(z)$. It is often assumed that the error in the radiosonde number density can be ignored. This assumption is tested below.

In the analysis to follow, two mathematically equivalent forms of the derivative function in Eq. (1) are used:

$$\frac{\mathrm{d}}{\mathrm{d}z}\left\{\ln\left[\frac{N(z)}{z^2P(z)}\right]\right\} = \tag{2}$$

$$\frac{1}{N(z)}\frac{\mathrm{d}}{\mathrm{d}z}N(z) - \frac{1}{z^2P(z)}\frac{\mathrm{d}}{\mathrm{d}z}[z^2P(z)]. \tag{3}$$

Various models are used to calculate the derivatives in expressions (2) and (3). The cumulative chi-squared (χ^2) distribution[4–6] is used to determine which equation and model combination is best suited to the set of data points being regressed. The most probable composite profile of aerosol extinction is determined here using these techniques. The results are compared with the standard technique of using a linear model to regress expression (2) and assigning no error to the number density data.

It should be noted that it is possible for one to avoid evaluating the numerical derivative mentioned here by calculating an average quantity of aerosol extinction or absorbing species within a layer of the atmosphere.[3] This can be done by integrating the derivative term over the extent of the layer. This is essentially an assumption of constant behavior of the quantity within the layer and linear behavior of the quantity between the layers. If one, however, desires to take advantage of the natural variation of the data both within and between layers, it is necessary to deal with the statistical issues outlined here.

3. Introduction to the Technique of Least-Squares Fitting

Following the treatment of Barlow,[4] consider a data sample consisting of a set of (x, y) pairs given by $[(x_i, y_i)]$. The x_i are assumed to be known exactly whereas the y_i have been measured each with some resolution σ_i. Suppose that the y_i are believed to be estimated by some function $f(x_i; \mathbf{a})$ where $\mathbf{a}$ is a vector of parameters that we want to estimate.

The central limit theorem states that the distribution of the measured y_i about their ideal values approaches a Gaussian. Thus the probability of a particular y_i for a given x_i is

$$P(y_i; \mathbf{a}) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\{-[y_i - f(x_i; \mathbf{a})]^2/(2\sigma_i^2)\}.$$

To maximize the likelihood that the $f(x_i; \mathbf{a})$ represent the y_i, one has to minimize the quantity

$$\chi^2 = \sum_{i=1}^{N}\frac{[y_i - f(x_i; \mathbf{a})]^2}{\sigma_i^2} = \sum_{i=1}^{N}\left(\frac{y_i^{\text{measured}} - y_i^{\text{ideal}}}{\text{measurement\_error}_i}\right)^2. \tag{4}$$

The minimization $\mathrm{d}\chi^2/\mathrm{d}\mathbf{a} = 0$ yields a set of simultaneous equations that when solved leads to the principle of least squares.

4. Comparison of Aerosol Extinction Calculations

The data used here were acquired by the NASA Goddard Space Flight Center Scanning Raman Lidar (SRL) system during the Convection and Moisture Experiment (CAMEX-II) that was held at Wallops Flight Facility during August and September 1995. The SRL uses a XeF excimer laser that produces an output spectrum centered at 351 nm. (The spectral output of the laser and the impact this has on measurements of Raman scattering in the atmosphere have been discussed elsewhere).[7] The backscattered light is collected by a 0.76-m-diameter Dahl–Kirkham telescope that is aligned with a 1.1 m × 0.8 m steerable mirror. This configuration allows data to be acquired from horizon to horizon in a single scan plane. Only vertical data are analyzed here, however. Through use of dichroic beam splitters and interference filters, the signal is split into four wavelengths: 351, 371, 382, and 403 nm. These signals are due to Rayleigh–Mie scattering, Raman scattering from oxygen, nitrogen, and water vapor, respectively. Each of these signals is split in approximately a 95% to 5% ratio and detected by a pair of photomultiplier tubes—one for the low-altitude return, the other for the high-altitude return—and finally recorded using 100-MHz photon counting. (The Rayleigh–Mie signal is attenuated by a factor of 10^{-3} to allow it to be photon counted.) The tropospheric aerosols studied here are confined to altitudes up to approximately 3 km. Because of this, only low-channel Rayleigh–Mie (351-nm) and Raman nitrogen (382-nm) data are used here. Although the SRL has daytime measurement capability, the data analyzed here were all acquired at night.

Data sets from 30 August, 3 September, and 13 September 1995 were examined. The data from 30 August and 3 September are averages over 5 min and the data from 13 September are averaged over 25 min. The data have a range resolution of 75 m. The results from the analysis of the 30 August data are presented here in detail. The analysis of the 3 and 13 September data yields similar conclusions.

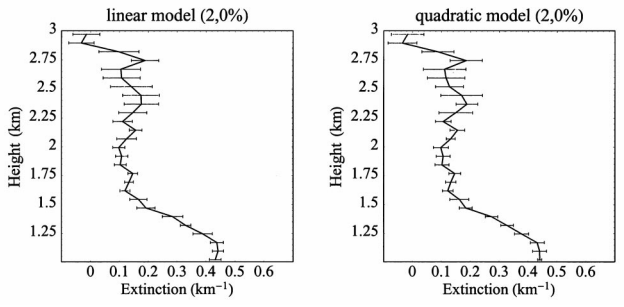

Fig. 1. Aerosol extinction calculated using both a linear and a quadratic model for the logarithm of the ratio in expression (2). The first plot shows the standard linear technique; the second plot shows a second-order polynomial fit of these same data. No error was assigned to the radiosonde number density data. The notation (2, 0%) indicates that expression (2) was regressed and that 0% error was assigned to the number density data.

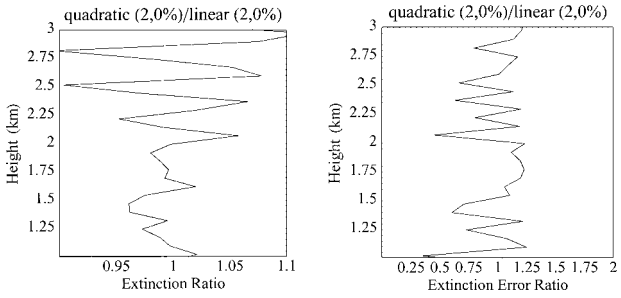

Fig. 2. Comparison of extinction values and extinction error of the data in Fig. 1. On the left is the ratio of the [extinction error using the quadratic regression for the derivative of expression (2)] and on the right is [the extinction error using the linear regression to determine the derivative of expression (2)]. The extinction values differ by up to ±10% at 3 km. However, the aerosol optical depth between 1 and 3 km agrees to better than 1% between the two techniques. On the right is shown the ratio of these extinction errors where the values can differ by up to 50% or more. On average, the error in determining the slope using the quadratic model is larger than the error in determining the slope using the linear model, with the quadratic model yielding an error in aerosol optical depth 14% higher than the linear model.

A. Example of using χ^2 to Select between a Linear and a Quadratic Model

Figure 1 shows two methods of calculating aerosol extinction using data acquired on 30 August 1995. The left plot uses a weighted linear regression to calculate the derivative in expression (2) using the standard technique mentioned above whereas the right plot uses a quadratic regression. [The notation in the plot titles indicates that, for example, in the left plot where (2, 0%) is noted, expression (2) is regressed using a linear model while using 0% error for the number density data. This convention is used for all plots.] Using a linear model, the slope of the best-fit line gives the derivative whereas the error in the determination of the derivative is given by the general first-order error propagation formula shown in Appendix A [Eq. (A1)] as follows. Using $g(z) = a + bz$ as the model for $\ln[N(z)/z^2P(z)]$, the desired derivative $g'(z)$ and its variance, letting $f(z) = g'(z)$, become

$$f(z) = g'(z) = b, \qquad \sigma_{f(z)}^{\ 2} = \sigma_b^{\ 2}. \tag{5}$$

So using a linear model, the error in determining the slope is given simply by σ_b.

The weights used in the linear regression of $\ln[N(z)/z^2P(z)]$ are determined by using Eq. (A1) to propagate errors and assuming that the error in the lidar data is given by Poisson statistics where $\sigma_i^{\ 2}$ is given by $P(z)$, the number of photon counts detected at height z. To calculate these weights, two specific formulas that result from applying the error propagation formula [Eq. (A1)] and assuming no covariance are needed. For $f(x) = \pm a(u/v)$, the variance is given by

$$\frac{\sigma_{f(x)}^{\ 2}}{f(x)^2} = \frac{\sigma_u^{\ 2}}{u^2} + \frac{\sigma_v^{\ 2}}{v^2},$$

and for $f(x) = a \ln(\pm bu)$, the variance is given by

$$\sigma_{f(x)}^{\ 2} = \left(a\,\frac{\sigma_u}{u}\right)^2.$$

Combining these for the case in which $g(z) = \ln[N(z)/z^2P(z)]$ results in

$$\sigma_{g(z)}^{\ 2} = \frac{\sigma_{P(z)}^{\ 2}}{P(z)^2} + \frac{\sigma_{N(z)}^{\ 2}}{N(z)^2}.$$

These are the weights that must be given in the minimization of Eq. (4). In the standard technique the error in the atmospheric density is ignored so that

$$\sigma_{g(z)}^{\ 2} = \frac{\sigma_{P(z)}^{\ 2}}{P(z)^2}.$$

On the right side of Fig. 1 is presented aerosol extinction using a quadratic model for $\ln[N(z)/z^2P(z)]$. In this case, using $g(z) = a + bz + cz^2$ as the model for $\ln[N(z)/z^2P(z)]$, the derivative and variance become

$$f(z) = g'(z) = b + 2cz,$$
$$\sigma_{f(z)}^{\ 2} = \sigma_b^{\ 2} + 4z^2\sigma_c^{\ 2} + 4z\sigma_{bc}^{\ 2}, \tag{6}$$

with the weights used in the regression being the same as in the linear case. In both cases (and in the cases that are considered below), a sliding window of five consecutive data points is used in the computation of aerosol extinction. The spacing of the resulting aerosol extinction values is 75 m; however, the resolution of the derived extinction values is 375 m. In the linear case on the left, the error bars plotted are given by the standard error of the slope given by the expression for $\sigma_{f(z)}^{\ 2}$ in Eqs. (5). In the quadratic case on the right, the standard error of the slope is given by the expression for $\sigma_{f(z)}^{\ 2}$ in Eqs. (6).

It has been claimed[2] that these two methods of performing the log derivative shown in Fig. 1 yield identical results. As can be seen in Fig. 2, however, this is only approximately true both in the case of

extinction value and in the case of extinction error. On the left side of Fig. 2, the ratio of the aerosol extinction computed using the quadratic model to that computed using the linear model is plotted. The difference between the two methods generally increases with height as the error in the lidar data increases. The difference in the ratios ranges from approximately ±5% between 1 and 2 km, increasing to ±10% at 3 km. The right side of Fig. 2 shows that the extinction errors as calculated by the two techniques differ by up to a factor of 2 at some altitudes. The integrated aerosol extinction, i.e., aerosol optical depth, between 1 and 3 km calculated using the two techniques agrees to better than 1% between the two methods. However, the error in aerosol optical depth is 14% larger using the quadratic technique.

To determine which of these two models is more probable, one can use the well-known χ^2 test. On average the minimum χ^2 value determined from a weighted least-squares regression will approximately equal the number of degrees of freedom in the regression. The number of degrees of freedom is given by $n - m$ where n is the number of data points in the regression (five in the analyses done here) and m is the number of parameters in the model (two in the case of a linear model, three for a quadratic model). Using the cumulative χ^2 distribution, the minimum value of χ^2 obtained from the regression can be used to calculate the probability that the combination of the model chosen and the errors attributed to the data (which determine the weights used in the regressions) accurately represent the data. The χ^2 probability distribution for ν degrees of freedom is given by

$$P_x(x^2; v) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu-2} \exp - \left(\frac{x^2}{2}\right).$$

The cumulative χ^2 distribution then is given by

$$P_\chi(\chi^2; v) = \int_{\chi^2}^{\infty} P_x(x^2; v)\mathrm{d}x^2. \quad (7)$$

A value of χ^2 such that Eq. (7) is approximately equal to 0.5 indicates that the combination of model and data are reasonable. This minimum χ^2 will then be close to the number of degrees of freedom. If the minimum value of χ^2 determined in the weighted regression greatly differs from the number of degrees of freedom in the regression, this is an indication of either an inappropriate choice of model or errors that do not accurately describe the actual variation in the data or both.

Figure 3 shows χ^2 and $P_\chi(\chi^2; v)$ values resulting from the regressions used to determine the aerosol extinction for both models used in Figs. 1 and 2. In the upper left panel of Fig. 3, the χ^2 values resulting from the linear regression are shown. Using a five-point regression to determine aerosol extinction leaves 3 degrees of freedom remaining. Thus a good model and data combination would yield minimum χ^2 values of approximately 3. The χ^2 values shown for the linear regression range from below 2 to approximately 10 with an average of 5.0. In the upper right panel of Fig. 3 is shown the χ^2 values resulting from the quadratic regression. The number of degrees of freedom in this case is 2 and χ^2 ranges from less than 1 to approximately 5 for most of the profile with an average of 2.7. The $P_\chi(\chi^2; v)$ values are shown in the lower panels.

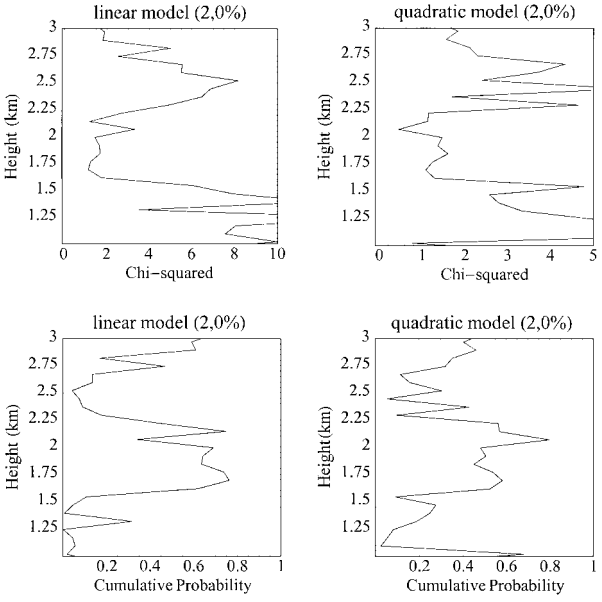

Fig. 3. χ^2 and cumulative probability plots for the regressions of Figs. 1 and 2. In general, a χ^2 minimum should approximately equal the degrees of freedom in a regression. Using the linear model on the left, the number of degrees of freedom is 3. Using the quadratic model on the right, the number of degrees of freedom is 2. A cumulative probability of approximately 0.5 indicates that a reasonable χ^2 minimum value was achieved.

The bottom panels of Fig. 3 show the cumulative χ^2 probability for both cases. A composite profile could be formed from these two profiles by choosing at each height the results from the model that has a cumulative χ^2 value closest to 0.5. However, before taking this step we will consider other methods to calculate aerosol extinction below. Prior to this, however, we must consider the influence of radiosonde number density errors on the calculation of aerosol extinction. In Subsection 4.B we demonstrate a technique for estimating the error in the number density data. This error is needed to perform a regression of expression (3).

In the analysis of this subsection, the technique of least-squares minimization is applied to the quantity $\ln[N(z)/z^2P(z)]$. Although both $N(z)$ and $z^2P(z)$ are taken to be normally distributed [at high count levels such as in $P(z)$ the Poisson distribution, which characterizes photon-counting data such as these, is essentially indistinguishable from a Gaussian distribution], their ratio is not, nor is the log of their ratio. Deriving a minimization technique for non-normally distributed data is beyond the scope or intent of this paper. What is desired, instead, is to examine the standard techniques that are in use in

the analysis of aerosol extinction and DIAL data and to demonstrate how statistical techniques can be used to improve these techniques. It should be noted that a regression of expression (3) avoids the difficulty of regressing non-Gaussian data because both the radiosonde data and the number density data are regressed separately maintaining their original statistical distributions. This different approach is examined in Subsection 4.D.

B. Assigning an Error to the Radiosonde Number Density

In the analysis of Figs. 1 and 2, the error in the radiosonde density measurements was ignored. However, the error that is attributed to these data affects both the weighting that occurs in a regression and the resultant values for both extinction and extinction error. Thus to fully apply the statistical techniques to be described below, an error must first be assigned to the radiosonde number density.

An estimate of the number density error can be obtained by using the χ^2 test in reverse, i.e., assume a certain model for the radiosonde number density and then choose the error that yields a value of χ^2 corresponding to a 0.5 cumulative probability. This technique requires one to make *a priori* decisions about which models to use to represent the data, however. These decisions can be further tested using the Run Test[4] that considers the sequence of residuals in the regression and determines how probable the sequence is. The run test was not used here because assigning an error to the radiosonde number density is not the main purpose of this study. Instead, what is intended here is to demonstrate (1) that one can use statistical techniques to estimate the error in data based on the natural variation that they display about some assumed model and (2) that the radiosonde number density error contributes to both the calculated extinction value and the estimate of extinction error.

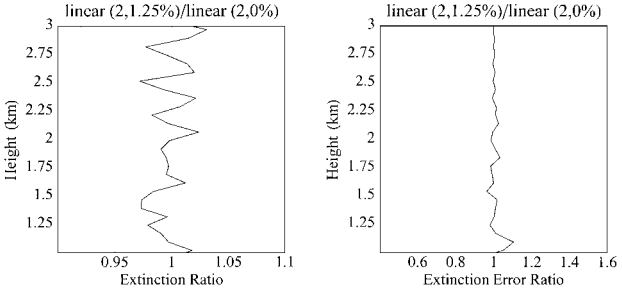

Fig. 4. Effect of the error attributed to radiosonde number density on the calculation of extinction and extinction error is shown here. On the left is the ratio of (extinction calculated with a linear model and assigning 1.25% to the number density error) and (the extinction calculated with a linear model and assigning 0% to the number density error). The two calculations differ by ±3% over the range of the profile. In the comparison of the extinction errors on the right, it can be seen that the effect of using errors in the number density are mainly confined to the lowest part of the profile where the two techniques differ by up to 10%. This is due to the fact that the lidar data errors are generally much smaller than the radiosonde error in the lowest part of the profile, but they are larger above 2 km and thus become dominant. Aerosol optical depth and error in aerosol optical depth agree to better than 1% between the two techniques.

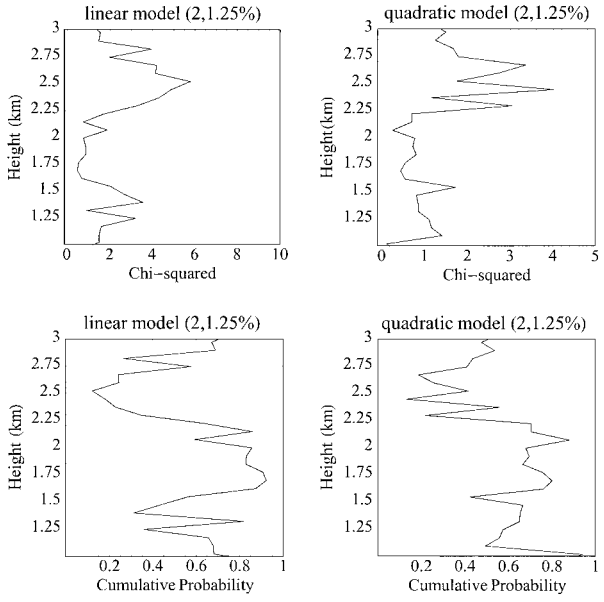

Fig. 5. Effect of using 1.25% number density error on χ^2 and cumulative χ^2 probability are shown here. Comparing these plots with those of Fig. 3 shows that including error for the number density yields more probable results between 1 and 2 km. Above 2 km where lidar error dominates the number density error, there is little difference between the two.

A value for the error in the radiosonde density data is now determined using expression (3). To calculate extinction with this equation, it is necessary to separately regress the radiosonde number density data $N(z)$ and the range-corrected lidar data $z^2P(z)$. The assumption is made here that over the restricted range of 1–3 km the radiosonde percentage error is constant. Using this assumption, the most probable radiosonde number density error can be determined with the cumulative χ^2 distribution to assess probability.

Use of a linear model to determine $d/dz[N(z)]$ in expression (3) yields a most probable value of approximately 1.75% error in density; i.e., this is the value that yields a minimum χ^2 equal to the 50% cumulative probability for 3 degrees of freedom (a value of approximately 2.4). Use of a quadratic model yields 1.25% for the radiosonde density error using the same technique. It is assumed here that because the atmospheric density data generally show an exponential decay with height, some curvature in the model is needed to best describe the data. Because of this, the quadratic model is chosen over the linear model to describe the radiosonde data. If the intent of the analysis here were to investigate radiosonde number density errors in detail, one could use the Run Test mentioned above to further assess the choice of model. That is not the intent here and thus the value of 1.25% is used for the error in the radiosonde density data in the subsequent analysis.

Figure 4 shows the effect of including a value of 1.25% for the error in radiosonde number density on both the calculated extinction and the extinction er-

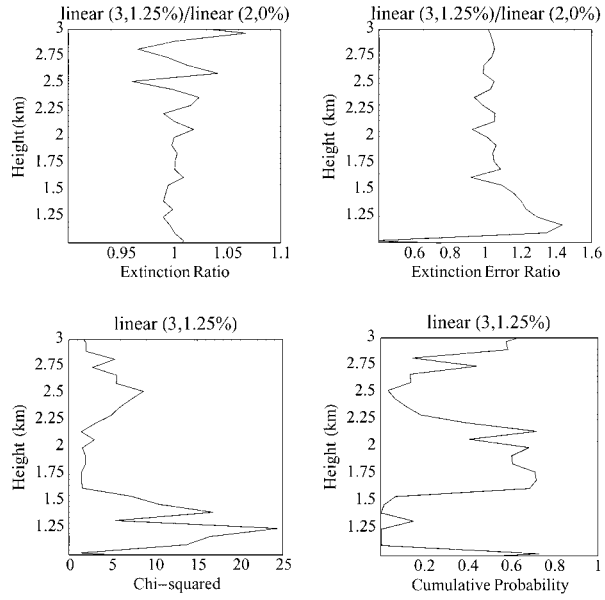

Fig. 6. Comparison of aerosol extinction and extinction error using a linear model on expression (3) to the standard technique. The extinction values differ by up to 5% in the upper portion of the profile. Extinction errors differ by up to 40% in the lower part of the profile. Aerosol optical depth calculated with the two techniques agrees to within 1% whereas the linear regression of expression (3) yields an error in aerosol optical depth that is 5% more than the standard technique. This is choice (3) for the composite model. Plots of χ^2 and cumulative χ^2 probability are shown for reference.

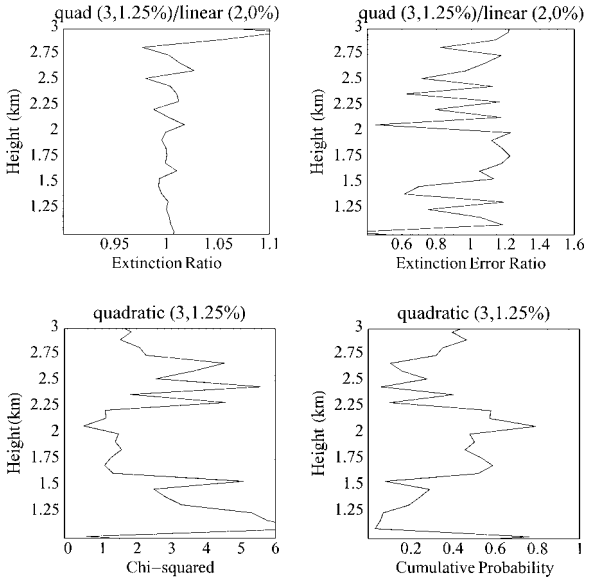

Fig. 7. Comparison of aerosol extinction and extinction error using a quadratic model on expression (3) to the standard technique. The extinction values differ by up to 10% in the upper portion of the profile whereas the errors differ by up to 50%. Aerosol optical depth calculated with the two techniques agrees to within 1% whereas the quadratic regression yields an error in aerosol optical depth that is 2% less than the standard technique. The quadratic model used on expression (3) is choice (4) for the composite model. Plots of χ^2 and cumulative χ^2 probability are shown for reference.

ror. As shown in the left plot, using 1.25% error for the radiosonde number density generally makes a $\pm 3\%$ difference in the value of extinction over the entire range of the extinction profile. The difference in the calculated errors ranges from $\pm 10\%$ at 1 km to near zero at 3 km reflecting the fact that the error in the lidar data is generally much smaller than the radiosonde error at 1 km and increases with altitude so that at 3 km the lidar error dominates the radiosonde number density error. Thus at 3 km the lidar error is the main determinant of the total error in extinction whether or not the errors are prescribed for the radiosonde density. The aerosol optical depth and the error in aerosol optical depth agree to within 1% between the two techniques.

Figure 5 shows the corresponding χ^2 and cumulative χ^2 plots for these regressions. Comparing Fig. 5 with Fig. 3 shows that including the 1.25% radiosonde error yields values for these quantities that are much more probable in the lowest 0.5 km. This is the region where including error for the number density significantly affects the calculated error.

C. Determining the Most Probable Composite Extinction Profile

The χ^2 test that was used in Subsection 4.B to determine the most probable error in the radiosonde number density can also be used to determine the most probable composite profile of aerosol extinction. In seeking the most probable composite profile for these aerosol extinction data, four choices for determining $d/dz\{\ln[N(z)/z^2P(z)]\}$ were considered: [1] performing a linear regression to determine the value of expression (2), [2] performing a quadratic regression to determine the value of expression (2), [3] performing a linear regression to determine the value of expression (3), and [4] performing a quadratic regression to determine the value of expression (3). For all these, 1.25% was used for the error in the number density. The equations pertaining to these techniques are given in Appendix A. Third-order models were also studied but gave much poorer results.

Choices (1) and (2) above are illustrated in Fig. 5. Figures 6 and 7 show the results using choices (3) and

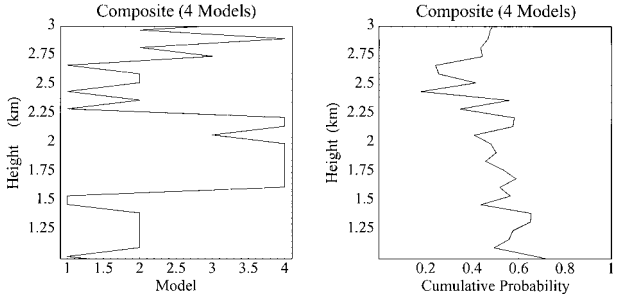

Fig. 8. Composite profile uses the results of one of four different models at each height. The model used is shown in the plot on the left whereas the total resulting cumulative χ^2 probability is shown on the right. Comparison with Figs. 5–7 shows the improvement of the cumulative probability over any of the individual models.

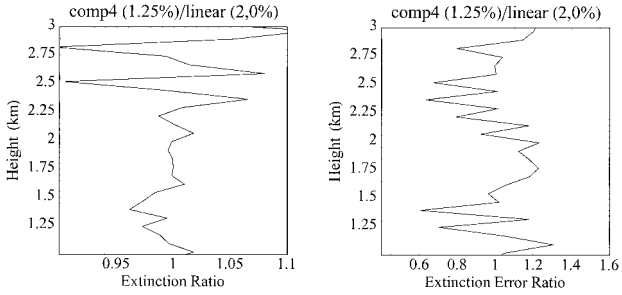

Fig. 9. Comparison of the four-model composite to the standard technique. The comparison of extinction is on the left and the comparison of extinction error is on the right. The extinction values differ by up to 10% at the top of the profile whereas the errors differ by up to 40%. Aerosol optical depth and error in aerosol optical depth agree to within 1% between the two techniques.

(4), respectively. In the upper plots are shown the extinction ratio and extinction error ratio compared with the standard technique [linear regression of expression (2) using 0% number density error]. The χ^2 and P_χ $(\chi^2; \upsilon)$ values are shown in the lower plots.

The most probable composite profile is now determined by comparing the results of the four methods of calculating extinction mentioned above. At each height, extinction and extinction error are calculated using all four methods. The cumulative χ^2 probability distribution is used at each height to determine which is the most probable model to use. The model that achieves a cumulative probability closest to 0.5 is chosen at that height. This procedure is repeated for all heights in the profile. Figure 8 shows the models that are used in the composite profile and the resulting $P_\chi(\chi^2; \upsilon)$. Note the improvement in cumulative χ^2 over any of the individual models.

Figure 9 shows the comparison of extinction and extinction error calculated using the composite profile and the standard technique. The values of extinction differ by up to 10% and the errors can differ by more than 50%. The aerosol optical depth and the error in aerosol optical depth agree to within 1% between the two techniques.

Figure 10 again shows a comparison of the most probable composite profile and the linear model, but here an error of 1.25% in radiosonde number density

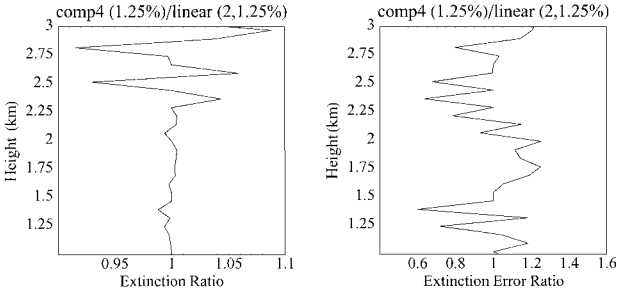

Fig. 10. Same as Fig. 9 except that 1.25% is used for number density error instead of 0%. The extinction values differ by up to 9% and the extinction errors differ by up to 40%. Aerosol optical depth and error in aerosol optical depth agree to within 1%.

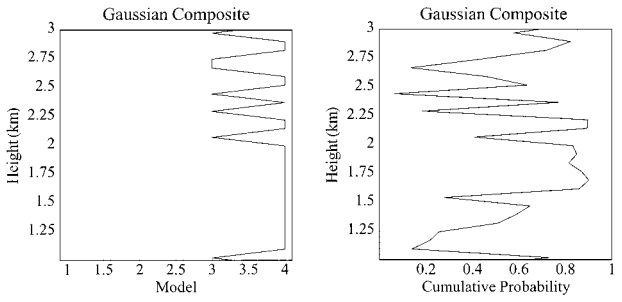

Fig. 11. Gaussian composite profile that uses choices (3) and (4) from the set of four models. The choice of model as a function of height and the cumulative χ^2 probability are shown.

is used in both. The calculated extinction values are now in better agreement, differing by up to 9% at 3 km, but the extinction errors still differ by 50% or more. The aerosol optical depth and the error in aerosol optical depth agree to within 1% between the two techniques.

D. Determining the Composite Profile by Regressing Only Normally Distributed Data

The preceding analysis was focused on comparing the usual method of calculating aerosol extinction with alternate methods that employ statistical analysis tests. The usual method, referred to as the standard technique, involves performing a linear regression of $\ln[N(z)/z^2P(z)]$ and assuming no error in the radiosonde number density values. The composite profile analyzed in Figs. 8, 9, and 10 uses linear and quadratic regressions of $\ln[N(z)/z^2P(z)]$ as two of the candidates for the final profile. However, as mentioned above, $\ln[N(z)/z^2P(z)]$ is not normally distributed. This violates one of the assumptions of the derivation of the least-squares technique. To assess the influence that this might have on the calculation of the most probable composite profile, this profile is now determined with only Gaussian-distributed data.

Expression (3) can be used to derive aerosol extinction in a way that allows all regressions to be performed on Gaussian-distributed data. The most probable composite profile has again been generated

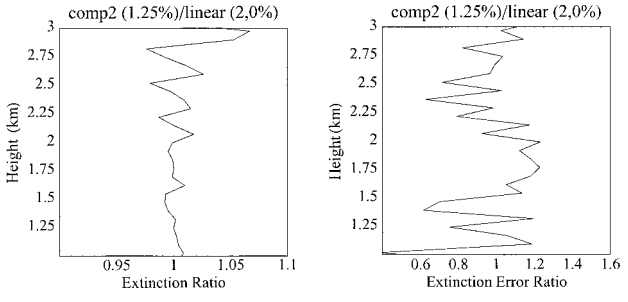

Fig. 12. Comparison of the two-model Gaussian composite profile to the standard technique. The techniques agree to within 5% in the calculation of extinction, but the errors differ by up to 60%. Aerosol optical depth calculated by the two techniques agrees to within 1%. The error in aerosol optical depth is 3% less using the two-model composite.

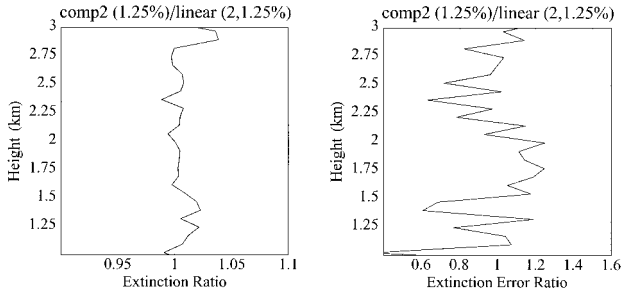

Fig. 13. Same as Fig. 12 except that 1.25% is used for the number density error instead of 0%. Differences between the two techniques are less than approximately 3% over the range of the profile. Errors differ by up to 40%. Aerosol optical depth calculated by the two techniques agrees to within 1%. The error in aerosol optical depth is 3% less using the two-model composite.

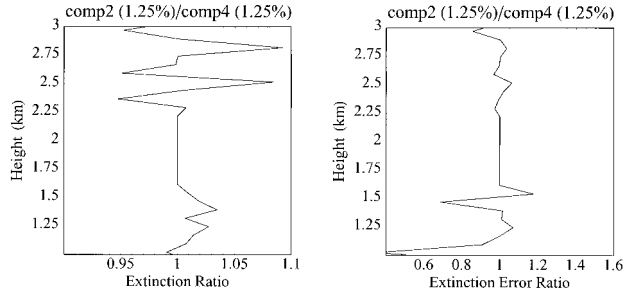

Fig. 14. Comparison of the four-model composite with the two-model Gaussian composite. The techniques differ by up to 8% at the upper end of the profile whereas the errors are in good agreement except in the lowest 0.5 km where they differ by up to 50%.

using the results of the cumulative χ^2 test used in Subsection 4.C. This time, however, linear and quadratic regressions of expression (2) are not candidates. The two remaining candidates that are considered at each height are the linear and quadratic models used to regress expression (3). The results are shown in Fig. 11 where the model used and the resulting cumulative probability are given as a function of height.

Figure 12 presents comparisons of the most probable aerosol extinction calculated from the two models described above to the standard technique. In the left plot, the ratio of the Gaussian composite to the standard technique is plotted. They agree to within approximately $\pm 5\%$. In the right plot, the ratio of the extinction errors for these profiles is plotted. Here the calculations disagree by up to 60%. The aerosol optical depth agrees between the two techniques to better than 1%; however, the error in aerosol optical depth is approximately 3% lower using the Gaussian composite.

Figure 13 shows the comparison of the composite profile versus the standard technique using 1.25% error for the radiosonde number density error. The extinction values now agree to within $\pm 4\%$. The errors still differ by up to 60%. For completeness, Fig. 14 shows the comparison of the most probable profile using all four candidate models and the most probable profile using just the normally distributed data sets. The calculated extinction differs by up to 2–3% in the first kilometer and up to 5% between 2–3 km. The extinction errors generally agree well except in the first 0.5 km where the difference is greater than 50% at 1 km.

5. Summary and Conclusions

In the calculation of aerosol extinction from Raman lidar data or of the concentration of an absorbing species using DIAL data, one must evaluate a numerical derivative. Various techniques for computing this derivative have been compared using aerosol extinction calculated from three nights of data acquired by the NASA Goddard Space Flight Center SRL during the CAMEX-II experiment held at Wallops Flight Facility in 1995. Statistical methods have also been used to estimate the radiosonde number density error. The same techniques were then used to determine the most probable model for the least-squares regression that gives the numerical derivative required to calculate extinction. The analysis shows that differences in calculated extinction as large as 10% are made using the standard technique versus the most probable composite profile determined by these statistical methods. Calculated extinction errors can differ by up to $\pm 40\%$ between the two methods of calculation.

The method of least-squares fitting assumes that the data are normally distributed. This assumption is violated for data of the type $\mathrm{d}/\mathrm{d}z[N(z)/z^2P(z)]$ as used in this study and as commonly used in DIAL analyses. It is possible, however, to reformulate the equations in a manner that allows the least-squares technique to be used on Gaussian-distributed data. It was shown that there is essentially no overall bias between the extinction results of this reformulated approach and those of the standard technique. However, the reformulated approach yields errors that are smaller by approximately 3% whether errors are attributed to the radiosonde number density.

Based on these results, if the desired accuracy in an extinction calculation at a given altitude is 1% or better, these statistical weighting issues must be taken into consideration. If calculations to an accuracy of 10% are sufficient, they can be disregarded. However, if the actual size of the error in the calculated extinction is of importance, the analysis here shows that use of statistical methods is required and that one should pay attention to whether the data being regressed have a Gaussian distribution. Reformulation of the equations in a manner that allows Gaussian data to be regressed is recommended.

Although no DIAL data were analyzed as a part of this study, the conclusions drawn here should pertain to DIAL calculations as well. The same statistical issues pertain to the calculation of the numerical derivative that is needed to determine the amount of the absorbing species. Also, DIAL measurements of stratospheric ozone, for example, can be of higher signal-to-noise ratio than these Raman measurements of aerosol extinction. Thus the significance of the differences in numerical derivative calculations

reported here are potentially of even greater concern in DIAL calculations.

Appendix A

1. Matrix Formulation of χ^2

The χ^2 function was given in the text as

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(z_i; \mathbf{a})]^2}{\sigma_i^2}.$$

If the σ_i measurement errors are correlated, this becomes a matrix equation,

$$\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{f}),$$

where $\mathbf{y} = \{y_i\}$, $\mathbf{f} = \{f(z_i; \mathbf{a})\}$, $\tilde{\mathbf{f}}$ is the transpose of $\mathbf{f}$, and the covariance matrix for the measurements $\mathbf{y}$ is $\mathbf{V}$. For example, in the case in which $\mathbf{a} = \{a, b, c\}$, $\mathbf{V}$ is given by

$$\mathbf{V} = \begin{pmatrix} \mathrm{var}(a) & \mathrm{cov}(ab) & \mathrm{cov}(ac) \\ \mathrm{cov}(ba) & \mathrm{var}(b) & \mathrm{cov}(bc) \\ \mathrm{cov}(ca) & \mathrm{cov}(cb) & \mathrm{var}(c) \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab}^2 & \sigma_{ac}^2 \\ \sigma_{ba}^2 & \sigma_b^2 & \sigma_{bc}^2 \\ \sigma_{ca}^2 & \sigma_{cb}^2 & \sigma_c^2 \end{pmatrix},$$

where it is noted that $\mathrm{cov}(ab) = \mathrm{cov}(ba)$, etc.

2. How the Measurement Errors Propagate into the Errors in the Derived Coefficients[4]

The treatment of Barlow[4] will be followed here. By differentiating χ^2 with respect to the a_i and setting these equal to zero, a set of simultaneous equations is derived. These can be solved for the vector $\hat{\mathbf{a}}$, which is the least-squares estimate of $\mathbf{a}$. If $f(z_i; \mathbf{a})$ is linear in the a_i, then the equations are linear and can be solved exactly.

Consider f at a particular z_0. For f a linear function, it can be represented as

$$f(z_0; \mathbf{a}) = \sum_r c_r(z) a_r.$$

$\mathbf{f}$ and χ^2 can now be represented in matrix form as

$$\mathbf{f} = \mathbf{C}\mathbf{a},$$

$$\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{a}}\tilde{\mathbf{C}})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{a}),$$

where $\mathbf{C}$ is an $n \times m$ matrix where n is the number of points z_i and m is number of coefficients c_r. Minimization of χ^2 with respect to $\mathbf{a}$ yields the best estimate of $\hat{\mathbf{a}}$ and the errors in the estimate $\mathbf{V}(\hat{\mathbf{a}})$:

$$\hat{\mathbf{a}} = (\tilde{\mathbf{C}}\mathbf{V}^{-1}\mathbf{C})^{-1}\tilde{\mathbf{C}}\mathbf{V}^{-1}\mathbf{y},$$

$$\mathbf{V}(\hat{\mathbf{a}}) = [(\tilde{\mathbf{C}}\mathbf{V}(\mathbf{y})^{-1}\mathbf{C})^{-1}],$$

where $\mathbf{V}(\mathbf{y})$ and $\mathbf{V}(\hat{\mathbf{a}})$ denote the variance matrices for the measurements and the results, respectively. Specific formulas can be derived from these equations as follows. In the case of a linear model $y = mz + b$ with the σ_i equal, the results are

$$\hat{m} = \frac{\overline{zy} - \overline{z}\,\overline{y}}{\overline{z^2} - \bar{z}^2},$$

$$V(\hat{m}) = \sigma_m^2 = \frac{\sigma^2}{N(\overline{z^2} - \bar{z}^2)},$$

$$\hat{b} = y - \hat{m}\bar{z},$$

$$V(\hat{b}) = \sigma_b^2 = \frac{\sigma^2 \overline{z^2}}{N(\overline{z^2} - \bar{z}^2)} = V(\hat{m})\overline{z^2}.$$

3. General Formula for Error Propagation

Defining $\mathbf{V_f}$ as the error matrix for $\mathbf{f}$ where $V_{ij} = \mathrm{cov}(x_i, x_j)$ and letting

$$G_{ki} = \left(\frac{\partial f_k}{\partial x_i}\right),$$

the general formula for error propagation becomes in matrix form,

$$\mathbf{V_f} = \mathbf{G}\mathbf{V_x}\tilde{\mathbf{G}}.$$

As an example, consider $f = f(a, b, c, z)$, the general first-order error propagation formula for a particular known value of z can be determined with

$$G = \left(\frac{\partial f(z)}{\partial a}, \frac{\partial f(z)}{\partial b}, \frac{\partial f(z)}{\partial c}\right).$$

Then

$$V_{f(z)} = \left(\frac{\partial f(z)}{\partial a}, \frac{\partial f(z)}{\partial b}, \frac{\partial f(z)}{\partial c}\right) \begin{pmatrix} \sigma_a^2 & \sigma_{ab}^2 & \sigma_{ac}^2 \\ \sigma_{ba}^2 & \sigma_b^2 & \sigma_{bc}^2 \\ \sigma_{ca}^2 & \sigma_{cb}^2 & \sigma_c^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f(z)}{\partial a} \\ \frac{\partial f(z)}{\partial b} \\ \frac{\partial f(z)}{\partial c} \end{pmatrix},$$

so that

$$\begin{aligned} V_{f(z)} &= \sigma_{f(z)}^2 \\ &= \sigma_a^2 \left[\frac{\partial f(z)}{\partial a}\right]^2 + \sigma_b^2 \left[\frac{\partial f(z)}{\partial b}\right]^2 + \sigma_c^2 \left[\frac{\partial f(z)}{\partial c}\right]^2 \\ &\quad + 2\sigma_{ab}^2 \left[\frac{\partial f(z)}{\partial a}\right]\left[\frac{\partial f(z)}{\partial b}\right] + 2\sigma_{ac}^2 \left[\frac{\partial f(z)}{\partial a}\right]\left[\frac{\partial f(z)}{\partial c}\right] \\ &\quad + 2\sigma_{bc}^2 \left[\frac{\partial f(z)}{\partial b}\right]\left[\frac{\partial f(z)}{\partial c}\right]. \end{aligned} \tag{A1}$$

Unless the regression is linear and is performed about the mean of the data points, the covariance terms in general will not be zero and must be included.

4. Specific Formulas for the Equations used to Form the Composite Profiles

By use of $g(z) = a + bz$ as the model for $\ln[N(z)/z^2P(z)]$, the desired derivative $g'(z)$ and its variance as given by Eq. (A1), letting $f(z) = g'(z)$ [this is choice (1) from Subsection 4.C], are

$$f(z) = g'(z) = b, \qquad \sigma_{f(z)}^2 = \sigma_b^2.$$

By use of $g(z) = a + bz + cz^2$ as the model for $\ln[N(z)/z^2P(z)]$, the derivative and variance become [choice (2)]:

$$f(z) = g'(z) = b + 2cz,$$

$$\sigma_{f(z)}^2 = \sigma_b^2 + 4z^2\sigma_c^2 + 4z\sigma_{bc}^2.$$

Similarly for expression (3) the linear model is achieved by modeling $N(z)$ by $g(z) = a + bz$ and $z^2P(z)$ by $h(z) = c + ez$. Then the formulas for the derivative and its variance become [choice (3)]:

$$f(z) = \frac{g'(z)}{g(z)} - \frac{h'(z)}{h(z)} = \frac{b}{a + bz} - \frac{e}{c + ez}$$

$$\sigma_{f(z)}^2 = \frac{1}{(a + bz)^4}(\sigma_a^2 b^2 + \sigma_b^2 a^2 - 2\sigma_{ab}^2 ab) + \frac{1}{(c + ez)^4}(\sigma_c^2 e^2 + \sigma_e^2 c^2 - 2\sigma_{ce}^2 ce).$$

Use of quadratic models for expression (3) such that $N(z) : g(z) = a + bz + cz^2$ and $z^2P(z) : h(z) = d + ez + fz^2$ yields for the derivative and its variance [choice (4)]:

$$f(z) = \frac{g'(z)}{g(z)} - \frac{h'(z)}{h(z)} = \frac{b + 2cz}{a + bz + cz^2} - \frac{e + 2fz}{d + ez + fz^2}$$

$$\sigma_{f(z)}^2 = \frac{1}{(a + bz + cz^2)^4}[\sigma_a^2(b + 2cz)^2 + \sigma_b^2(a - cz^2)^2 + \sigma_c^2(2a + bz)^2 z^2 - 2\sigma_{ab}^2(b + 2cz)(a - cz^2) - 2\sigma_{ac}^2(2a + bz)(b + 2cz)z + 2\sigma_{bc}^2(2a + bz) \times (a - cz^2)z] + \frac{1}{(d + ez + fz^2)^4}[\sigma_d^2(e + 2fz)^2 + \sigma_e^2(d - fz^2)^2 + \sigma_f^2(2d + ez)^2 z^2 - 2\sigma_{de}^2 \times (e + 2fz)(d - fz^2) - 2\sigma_{df}^2(2d + ez)(e + 2fz)z + 2\sigma_{ef}^2(2d + ez)(d - fz^2)z].$$

Thanks go to Bimal Sinha and Philip Rous, both of the University of Maryland Baltimore County, for insightful discussions during the preparation of this manuscript. This effort was supported by NASA's Atmospheric Dynamics and Remote Sensing Program.

References

1. A. Ansmann, M. Riebesell, and C. Weitkamp, "Measurement of atmospheric aerosol extinction profiles with a Raman lidar," Opt. Lett. **15,** 746–748 (1990).
2. T. J. McGee, D. Whiteman, R. Ferrare, J. Butler, and J. Burris, "STROZ LITE: stratospheric ozone lidar trailer experiment," Opt. Eng. **30,** 31–39 (1991).
3. R. Measures, *Laser Remote Sensing Fundamentals and Applications* (Wiley-Interscience, London, 1984).
4. R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley, New York, 1989).
5. R. V. Hogg and E. A. Tanis, *Probability and Statistical Inference*, 4th ed. (MacMillan, New York, 1993).
6. J. R. Taylor, *An Introduction to Error Analysis—The Study of Uncertainties in Physical Measurements* (University Science, Mill Valley, Calif., 1982).
7. D. N. Whiteman, W. F. Murphy, N. W. Walsh, and K. D. Evans, "Temperature sensitivity of an atmospheric Raman lidar system based on a XeF excimer laser," Opt. Lett. **18,** 247–249 (1993).